

	Type	Hits	Search Text
1	BRS	128	extract\$3 with common with (metadata or attribut\$2)
2	BRS	38	S1 and "707"/\$.ccls.
3	BRS	23	(extract\$3 with common with (metadata or attribut\$2)) same ((generat\$3 or creat\$3) with (metadat or attribut\$2))
4	BRS	0	S3 and ((attach\$3 or append\$3) with (director\$3 or folder\$1))
5	BRS	50	(extract\$3 with common with (metadata or attribut\$2)) and ((generat\$3 or creat\$3) with (metadat or attribut\$2))
6	BRS	0	S5 and ((attach\$3 or append\$3) with (director\$3 or folder\$1))
7	BRS	2	"6009439".pn.
8	BRS	0	(extract\$3 with (common with attribut\$2)) and (append\$3 with director\$3)
9	BRS	9	(extract\$3 with (meta-data or attribut\$2)) and (append\$3 with director\$3)
10	BRS	2	(extract\$3 with (meta-data or attribut\$2)) same ((append\$3 or attach\$3) with director\$3)
11	BRS	11	S11 and "707"/\$.ccls.
12	BRS	36	(extract\$3 with (meta-data or attribut\$2)) and ((append\$3 or attach\$3) with director\$3)
13	BRS	431	(extract\$3 with content\$1) same metadata
14	BRS	15	(extract\$3 with content\$1) same metadata same (directory\$3 or folder\$1)

	Type	Hits	Search Text
15	BRS	24	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1)) same metadata same (directory\$3 or folder\$1)
16	BRS	0	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1)) same metadata same (directory\$3 or folder\$1) same (attach\$3 or append\$3)
17	BRS	27	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1)) same metadata same (attach\$3 or append\$3)
18	BRS	191	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1) with (file\$1 or folder\$1)) same (attach\$3 or append\$3)
19	BRS	193	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1) with (file\$1 or folder\$1 or director\$3)) same (attach\$3 or append\$3)
20	BRS	2	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1) with (file\$1 or folder\$1 or director\$3)) same (attach\$3 or append\$3) same metadata
21	BRS	2	(extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or content\$1) with (document\$1 or file\$1 or folder\$1 or director\$3)) same (attach\$3 or append\$3) same metadata
22	BRS	101	((attach\$3 or append\$3) with (extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or metadata) with (document\$1 or file\$1 or folder\$1 or director\$3)))
23	BRS	17	S22 and "707"/\$.ccls.
24	BRS	16	((attach\$3 or append\$3) with (extract\$3 with (keyword\$1 or name\$1 or id or identifier\$1 or hyperlink\$1 or metadata))) same (director\$3 or folder\$1)

	Type	Hits	Search Text
25	BRS	188	(mp3 or MP3) and ID3
26	BRS	18	S25 and (@ad<"19990427" or @rlad<"19990427")
27	BRS	11	S25 and (@ad<"19990427" or @rlad<"19990427") and (classif\$5 or group\$3 or sort\$3)
28	BRS	6	S28 and "707"/\$.ccls.
29	BRS	30	S22 and (@ad<"19990427" or @rlad<"19990427")
30	BRS	7	((attach\$3 or append\$3) with (extract\$3 with (keyword\$1 or name\$1 or hyperlink\$1 or metadata) with content with (document\$1 or file\$1 or folder\$1 or director\$3)))
31	BRS	92	((generat\$3 or creat\$3) with (common near2 (metadata or keyword\$1 or name\$1 or attribute\$1))) same (folder\$1 or director\$3 or storage\$1)
32	BRS	79	S31 and (group\$4 or classif\$5 or sort\$3)
33	BRS	24	S32 and "707"/\$.ccls.
34	BRS	3	S33 and @rlad<"19990427"
35	BRS	4	S32 and @rlad<"19990427"
36	BRS	457	((group\$4 or classif\$5 or sort\$3) with (metadata or attribute\$1 or name\$1 or keyword\$1 or (file near2 extension\$1))) same ((generat\$3 or creat\$3) with (director\$3 or folder\$1))

	Type	Hits	Search Text
37	BRS	15	((group\$4 or classif\$5 or sort\$3) with (common\$5 or similar\$5) with (metadata or attribute\$1 or name\$1 or keyword\$1 or (file near2 extension\$1))) same ((generat\$3 or creat\$3) with (director\$3 or folder\$1))
38	BRS	2	S37 and @rlad<"19990427"
39	BRS	112	((generat\$3 or creat\$3) with (director\$3 or folder\$1)) same (common\$4 or similar\$5) near2 (name\$1 or attribute\$1 or metadata or keyword\$1 or ((file or folder or director\$3) near extension\$1))
40	BRS	112	((generat\$3 or creat\$3) with (director\$3 or folder\$1)) same ((common\$4 or similar\$5) near2 (name\$1 or attribute\$1 or metadata or keyword\$1 or ((file or folder or director\$3) near extension\$1)))
41	BRS	22	S40 and @rlad<"19990427"
42	BRS	2	(extract\$3 with common with (metadata or name\$1 or id or identifier\$1)) same ((generat\$3 or creat\$3) with (director\$3 or folder\$1 or categor\$3))
43	BRS	2	S42 and 707/1,10,101,104.1.ccls.
44	BRS	2	S42 and 707/1,10,101,104.1,103y,102z.ccls.
45	BRS	2	(extract\$3 with common with (metadata or name\$1 or id or identifier\$1)) same ((generat\$3 or creat\$3) with (director\$3 or folder\$1 or categor\$3))
46	BRS	2	S45 and 707/1,10,101,104.1,103y,103z.ccls.


[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☐ The Guide

 **SEARCH**

PREVIOUS WORKS BY AUTHOR


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

## TA-RE: an exchange language for mining software repositories

Full text Pdf (287 KB)

**Source** [International Conference on Software Engineering archive](#)  
**Proceedings of the 2006 international workshop on Mining software repositories** [table of contents](#)  
 Shanghai, China  
**SESSION: Repositories** [table of contents](#)  
 Pages: 22 - 25  
 Year of Publication: 2006  
 ISBN:1-59593-397-2

<b>Authors</b>	<a href="#">Sunghun Kim</a>	University of California, Santa Cruz, CA
	<a href="#">Thomas Zimmermann</a>	Saarland University, Saarbrücken, Germany
	<a href="#">Miryung Kim</a>	University of Washington
	<a href="#">Ahmed Hassan</a>	University of Waterloo, Canada
	<a href="#">Audris Mockus</a>	Avaya labs
	<a href="#">Tudor Girba</a>	University of Berne, Switzerland
	<a href="#">Martin Pinzger</a>	University of Zurich, Switzerland
	<a href="#">E. James Whitehead, Jr.</a>	University of California, Santa Cruz, CA
<a href="#">Andreas Zeller</a>	Saarland University, Saarbrücken, Germany	

**Sponsors** [ACM: Association for Computing Machinery](#)  
[SIGSOFT: ACM Special Interest Group on Software Engineering](#)

**Publisher** [ACM Press](#) New York, NY, USA

**Additional Information:** [abstract](#) [references](#) [index terms](#) [collaborative colleagues](#)

**Tools and Actions:** [Find similar Articles](#) [Review this Article](#)  
[Save this Article to a Binder](#) Display Formats: [BibTex](#) [EndNote](#) [ACM Ref](#)

**DOI Bookmark:** Use this link to bookmark this Article: <http://doi.acm.org/10.1145/1137983.1137990>  
[What is a DOI?](#)

### ↑ ABSTRACT

Software repositories have been getting a lot of attention from researchers in recent years. In order to analyze software repositories, it is necessary to first extract raw data from the version control and problem tracking systems. This poses two challenges: (1) extraction requires a non-trivial effort, and (2) the results depend on the heuristics used during extraction. These challenges burden researchers that are new to the community and make it difficult to benchmark software repository mining since it is almost impossible to reproduce experiments done by another team. In this paper we present the TA-RE corpus. TA-RE collects extracted data from software repositories in order to build a collection of projects that will simplify extraction process. Additionally the collection can be used for benchmarking. As the first step we propose an exchange language capable of making sharing and reusing data as simple as possible.

### ↑ REFERENCES

Note: OCR errors may be found in this Reference List extracted from the full text article. ACM has opted to expose the complete List rather than only correct and linked references.

- 1 J. Bevan and E. J. Whitehead, Jr., "Identification of Software Instabilities," Proc. of 2003 Working Conference on Reverse Engineering (WCRE 2003), Victoria, Canada, 2003.
- 2 J. Bevan, E. J. Whitehead, Jr., S. Kim, and M. Godfrey, "Facilitating Software Evolution with Kenyon," Proc. of the 2005 European Software Engineering Conference and 2005 Foundations of Software Engineering (ESEC/FSE 2005), Lisbon, Portugal, pp. 177--186, 2005.
- 3 D. Beyer and A. Noack, "Clustering Software Artifacts Based on Frequent Common Changes," Proc. of the 13th IEEE International Workshop on Program Comprehension (IWPC 2005), St. Louis, Missouri, USA, pp. 259--268, 2005.
- 4 V. Dallmeier, P. Weißgerber, and T. Zimmermann, "APFEL: A Preprocessing Framework For Eclipse," 2005, <http://www.st.cs.uni-sb.de/softevo/apfel/>.
- 5 S. G. Eick, T. L. Graves, A. F. Karr, J. S. Marron, and A. Mockus, "Does Code Decay? Assessing the Evidence from Change Management Data," IEEE Transactions on Software Engineering, vol. 27, pp. 1--12., 2001.
- 6 M. Fischer, M. Pinzger, and H. Gall, "Populating a Release History Database from Version Control and Bug Tracking Systems," Proc. of 2003 Int'l Conference on Software Maintenance (ICSM'03), pp. 23--32, 2003.
- 7 M. W. Godfrey and L. Zou, "Using Origin Analysis to Detect Merging and Splitting of Source Code Entities," IEEE Trans. on Software Engineering, vol. 31, pp. 166--181, 2005.
- 8 T. L. Graves, A. F. Karr, J. S. Marron, and H. Siy, "Predicting Fault Incidence Using Software Change History," IEEE Transactions on Software Engineering, vol. 26, pp. 653--661, 2000.
- 9 T. L. Graves and A. Mockus, "Inferring Change Effort from Configuration Management Data," Proc. of In Metrics 98: Fifth International Symposium on Software Metrics, Bethesda, Maryland, pp. 267--273, 1998.
- 10 M. Kim, V. Sazawal, D. Notkin, and G. Murphy, "An Empirical Study of Code Clone Genealogies," Proc. of the 2005 European Software Engineering Conference and 2005 Foundations of Software Engineering (ESEC/FSE 2005), Lisbon, Portugal, pp. 187--196, 2005.
- 11 S. Kim, K. Pan, and E. J. Whitehead, Jr., "When Functions Change Their Names: Automatic Detection of Origin Relationships," Proc. of 12th Working Conference on Reverse Engineering (WCRE 2005), Pennsylvania, USA, 2005.
- 12 S. Kim, E. J. Whitehead, Jr., and J. Bevan, "Analysis of Signature Change Patterns," Proc. of Int'l Workshop on Mining Software Repositories (MSR 2005), Saint Louis, Missouri, USA, pp. 64--68, 2005.
- 13 D. Lewis, Y. Yang, T. Rose, and F. Li, "RCV1: A New Benchmark Collection for Text Categorization Research " Journal of Machine Learning Research, vol. 5, pp. 361--397, 2004.
- 14 A. Mockus, R. F. Fielding, and J. Herbsleb, "A Case Study of Open Source Development: The Apache Server," Proc. of 22nd Int'l Conference on Software Engineering (ICSE 2000), Limerick, Ireland, pp. 263--272 2000.
- 15 A. Mockus and J. Herbsleb, "Expertise Browser: A Quantitative Approach to Identifying

Expertise," Proc. of 24rd Int'l Conference on Software Engineering (ICSE 2002), Orlando, Florida, pp. 503--512, 2002.

16 A. Mockus and L. G. Votta, "Identifying Reasons for Software Changes Using Historic Databases," Proc. of International Conference on Software Maintenance (ICSM 2000), San Jose, California, USA, pp. 120--130, 2000.

17 A. Mockus and D. M. Weiss, "Globalization by Chunking: a Quantitative Approach," IEEE Software, vol. 18, pp. 30--37, 2001.

18 A. Mockus, P. Zhang, and P. Li, "Drivers for Customer Perceived Software Quality," Proc. of 2005 Int'l Conference on Software Engineering (ICSE 2005), Saint Louis, Missouri, USA, 2005.

19 D. J. Newman, S. Hettich, C. L. Blake, and C. J. Merz, "UCI Repository of machine learning databases," 1988, <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

20 J. Sayyad Shirabad and T. J. Menzies, "The PROMISE Repository of Software Engineering Databases," 2005, <http://promise.site.uottawa.ca/SERepository>.

21 J. Sliwerski, T. Zimmermann, and A. Zeller, "When Do Changes Induce Fixes?" Proc. of Int'l Workshop on Mining Software Repositories (MSR 2005), Saint Louis, Missouri, USA, pp. 24--28, 2005.

22 T. Zimmermann and P. Weißgerber, "Preprocessing CVS Data for Fine-Grained Analysis," Proc. of Int'l Workshop on Mining Software Repositories (MSR 2004), Edinburgh, Scotland, pp. 2--6, 2004.

23 T. Zimmermann, P. Weißgerber, S. Diehl, and A. Zeller, "Mining Version Histories to Guide Software Changes," IEEE Trans. Software Engineering, vol. 31, pp. 429--445, 2005.

## ↑ INDEX TERMS

### Primary Classification:

**D.** Software

↳ **D.2** SOFTWARE ENGINEERING

↳ **D.2.7** Distribution, Maintenance, and Enhancement

↳ **Subjects:** Restructuring, reverse engineering, and reengineering

### Additional Classification:

**K.** Computing Milieux

↳ **K.6** MANAGEMENT OF COMPUTING AND INFORMATION SYSTEMS

↳ **K.6.3** Software Management

↳ **Subjects:** Software maintenance

### General Terms:

Experimentation, Measurement

### Keywords:

analysis, corpus, prediction, software repository mining

↑ **Collaborative Colleagues:**

<u>Tudor Girba:</u>	<u>Serge Demeyer</u> <u>Bart Du Bois</u> <u>Stephane Ducasse</u> <u>Mohammad El-Ramly</u> <u>Harald Gall</u> <u>Orla Greevy</u> <u>Dirk Janssens</u> <u>Adrian Kuhn</u> <u>Michele Lanza</u> <u>Cristina Marinescu</u>	<u>Radu Marinescu</u> <u>Tom Mens</u> <u>Jacek Ratzinger</u> <u>Matthias Rieger</u> <u>Filip Van</u> <u>Rysselberghe</u> <u>Mauricio Seeberger</u>	
<u>Ahmed Hassan:</u>	<u>Ahmed Abdel-Rahim</u> <u>Walid Abdelmoez</u> <u>Hany H. Ammar</u> <u>Kalaivani Appukkutty</u> <u>Vittorio Cortellessa</u> <u>Rania Elnaggar</u> <u>Katerina Goseva-Popstojanova</u> <u>Ajith R. Guedem</u> <u>Sherif Kamel</u> <u>Phillip Rust</u>		
<u>Miryung Kim:</u>	<u>Lawrence Bergman</u> <u>Tessa Lau</u> <u>Gail Murphy</u> <u>David Notkin</u> <u>Vibha Sazawal</u>		
<u>Sunghun Kim:</u>	<u>Jennifer Bevan</u> <u>Michael Godfrey</u> <u>Kai Pan</u> <u>Mark Slater</u> <u>E. James Whitehead</u> <u>E. James Whitehead</u>		
<u>Audris Mockus:</u>	<u>David Atkins</u> <u>David L. Atkins</u> <u>Thomas Ball</u> <u>William F. Eddy</u> <u>Stephen G. Eick</u> <u>Roy T. Fielding</u> <u>Thomas A. Finholt</u> <u>Birgit Geppert</u> <u>Todd Graves</u> <u>Todd L. Graves</u>	<u>Rebecca E. Grinter</u> <u>Ahmed E. Hassan</u> <u>James Herbsleb</u> <u>James D. Herbsleb</u> <u>Stacie Hibino</u> <u>Richard C. Holt</u> <u>Philip M. Johnson</u> <u>Alan F. Karr</u> <u>Mayuram Krishnan</u> <u>Paul Luo Li</u>	<u>J. S. Marron</u> <u>Shingo Oue</u> <u>Adam Porter</u> <u>Adam A. Porter</u> <u>Frank Rossler</u> <u>Harvey Siy</u> <u>Harvey P. Siy</u> <u>George T.</u> <u>Tucker</u> <u>Lawrence Votta</u> <u>Lawrence G.</u> <u>Votta</u>
<u>Martin Pinzger:</u>	<u>Jose L. Arciniegas</u> <u>Stéphane Ducasse</u> <u>Michael Fischer</u> <u>Beat Fluri</u> <u>Harald Gall</u> <u>Harald C. Gall</u> <u>Dharmalingam Ganesan</u> <u>Thomas Gschwind</u> <u>Mehdi Jazayeri</u> <u>Isabel John</u>	<u>Jens Knodel</u> <u>Michele Lanza</u> <u>Johann Oberleitner</u> <u>Claudio Riva</u> <u>Fernando Usero</u>	<u>David M. Weiss</u> <u>Ping Zhang</u>



<u>E. James</u> <u>Whitehead:</u>	<u>Jennifer Bevan</u> <u>Guozheng Ge</u> <u>Michael Godfrey</u> <u>Yaron Y. Golan</u> <u>Sunghun Kim</u> <u>Kai Pan</u> <u>Mark Slater</u> <u>Meredith Wiggins</u>	
<u>Andreas Zeller:</u>	<u>Jacek Śliwerski</u> <u>Jacek Śliwerski</u> <u>Philipp Bouillon</u> <u>Martin Burger</u> <u>Jong-Deok Choi</u> <u>Holger Cleve</u> <u>Valentin Dallmeier</u> <u>Stephan Diehl</u> <u>Cormac Flanagan</u> <u>Konstantin Halachev</u>	<u>Ralf Hildebrandt</u> <u>Dorothea</u> <u>Lütkehaus</u> <u>Karsten Lehmann</u> <u>Christian Lindig</u> <u>Gregor Snelting</u> <u>Peter Weisgerber</u> <u>Peter Weissgerber</u> <u>Thomas</u> <u>Zimmermann</u>
<u>Thomas</u> <u>Zimmermann:</u>	<u>Jacek Śliwerski</u> <u>Jacek Śliwerski</u> <u>Patricia Bomme</u> <u>Stéphane Commend</u> <u>Valentin Dallmeier</u> <u>Stephan Diehl</u> <u>Yves Dubois-Pèlerin</u> <u>Dominique Eyheramendy</u> <u>Konstantin Halachev</u> <u>Benjamin Livshits</u>	<u>Peter Weisgerber</u> <u>Peter Weissgerber</u> <u>Andreas Zeller</u>

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2006 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads:  [Adobe Acrobat](#)  [QuickTime](#)  [Windows Media Player](#)  [Real Player](#)



Welcome United States Patent and Trademark Office

[Search Session History](#)[BROWSE](#)[SEARCH](#)[IEEE XPLORE GUIDE](#)

Edit an existing query or  
compose a new query in the  
Search Query Display.

Mon, 7 Aug 2006, 8:40:27 AM EST

Search Query Display

Select a search number (#)  
to:

- Add a query to the Search Query Display
- Combine search queries using AND, OR, or NOT
- Delete a search
- Run a search

Recent Search Queries

- #1 ((extracting and common and (metadata or name or id or identification) and (generating or creating) and (directories or folders) )<in>metadata)
- #2 ((extracting and common and (metadata or name or id or identification) and (generating or creating) and (directories or folders) )<in>metadata)

Indexed by  
 Inspec

[Help](#) [Contact Us](#) [Privacy &](#)

© Copyright 2006 IEEE –